



Federated Learning-based Algorithm Design for Privacy Preservation in Cross-domain Data Sharing

Yuxin Wu

College of Engineering, Carnegie Mellon University, Moffett Field, CA 94035, USA.

How to cite this paper: Yuxin Wu. (2026) Federated Learning-based Algorithm Design for Privacy Preservation in Cross-domain Data Sharing. *Engineering Advances*, 6(1), 36-40.
DOI: 10.26855/ea.2026.03.008

Received: January 9, 2026
Accepted: February 6, 2026
Published: March 2, 2026

***Corresponding author:** Yuxin Wu, College of Engineering, Carnegie Mellon University, Moffett Field, CA 94035, USA.

Abstract

In the context of accelerating digital integration across industries, cross-domain data sharing has emerged as a critical enabler of intelligent decision-making and resource coordination. However, the sensitivity of data and disputes over ownership have made privacy protection the primary obstacle in collaborative environments. Federated learning, as a promising paradigm for distributed collaborative modeling, offers the key advantage of “data usability without visibility,” enabling privacy-aware computation. This study addresses major challenges—including data heterogeneity, regulatory compliance, and security threats—by proposing a privacy-enhanced federated algorithm framework. The design integrates multiple mechanisms such as dynamic aggregation, local perturbation, consistency regularization, and multi-layered security protection. Through this integrated approach, the framework enhances both the efficiency and security of federated models in heterogeneous scenarios, paving the way for a data-sharing ecosystem that safeguards privacy while enabling collaborative intelligence across critical domains.

Keywords

Federated Learning Privacy protection; Cross-domain collaboration Algorithm design Security mechanism

As “data silos” impose constraints on the intelligent evolution of industries, cross-domain data collaboration has gradually become the core proposition for the digital upgrade of various industries. Especially in highly sensitive scenarios such as healthcare, finance, and government affairs, the demand for collaboration among different data ownership subjects is constantly rising, while also exposing risks and hidden dangers such as data abuse, unclear boundaries, and algorithmic discrimination. The traditional centralized learning mechanism relies on data aggregation. While improving the modeling efficiency, it is difficult to balance the privacy rights and data control of the participants. As a result, the federated learning mechanism characterized by “local data retention and model sharing and update” has come into view. Under the premise of retaining data sovereignty, it realizes distributed intelligent modeling and has become an important means to break through the bottleneck of data sharing. The current key issue lies not only in “whether it can be federated”, but also in “how to connect safely and effectively”.

1. Privacy Protection Challenges in Cross-Domain Data Sharing

1.1 Analysis of the Difficulties in Cross-Domain Heterogeneous Data Collaboration

Against the backdrop of increasingly frequent data integration across multiple industries, the primary challenge for cross-domain data collaboration is not technical access, but the high heterogeneity of the underlying structure. Data from different fields show significant differences in terms of source, format, storage structure, and interface

standards. For instance, electronic medical records in the medical system are often dominated by unstructured text, while in the financial system, strongly structured transaction forms are at the core. The two are incompatible in feature space, timestamp mechanism, and even data density. This heterogeneity makes it difficult to unify the input and output formats in model training, seriously affecting the collaborative efficiency. Meanwhile, different data owners are subject to different legal frameworks. For instance, the scope of application of regulations such as GDPR and HIPAA varies by industry and region, making it difficult to clarify the boundaries of sharing. In addition, semantic inconsistency and label definition deviation further increase the difficulty of alignment [1]. Without a unified semantic mapping mechanism and privacy isolation structure to support cross-domain collaboration, it often falls into the predicament of “having the willingness to share but lacking the ability to share”.

1.2 Privacy Exposure Risks of Centralized Learning Models

The traditional centralized modeling path relies on the aggregation of raw data to a center and the completion of training on a unified platform [2]. Although this mechanism had certain efficiency advantages in the early model development, it exposed obvious privacy and security risks in multi-source sensitive data scenarios. Firstly, data is highly likely to become a target of attack during transmission, especially when there is a lack of encrypted links and authentication mechanisms; the risk of data hijacking and theft significantly increases. Secondly, even if the data aggregation is completed, the central server itself is a single-point risk core. Once it is attacked, it will cause systemic privacy leakage. Furthermore, research indicates that there is a possibility that the model parameters obtained through centralized training can be reverse-derived from the original data, especially in the absence of differential privacy protection. Malicious individuals can then achieve reverse identification of the data through means such as model inversion and member reasoning. In the current compliance environment that emphasizes “data sovereignty” and the “principle of minimum exposure”, the centralized model has become difficult to meet the data governance requirements of highly sensitive scenarios.

1.3 Compliance and Trust Thresholds in the Sharing Mechanism

In the practice of cross-domain data sharing, even with a foundation in collaborative technology, there are still severe trust and compliance thresholds in reality. On the one hand, data owners are usually concerned that once the data goes out of the domain, the purpose of use will be uncontrollable and the tracking chain will break, which is highly likely to cause problems such as unclear rights and responsibilities and risk spillover, especially when it involves personal privacy or business-sensitive information. On the other hand, the existing data sharing mechanisms generally lack full-process record-keeping and auditability mechanisms, making it difficult to trace and supervise data access behaviors, which increases the compliance burden and psychological resistance of the sharing parties. Furthermore, if there is no clear division of data responsibilities and a collaboration framework among the platform, users, and regulators, it is very likely to fall into a vicious cycle of “information asymmetry - trust deficiency - cooperation stagnation”, resulting in a structural gap between data utilization capabilities and governance mechanisms. Therefore, privacy protection is no longer merely a matter of encryption and anonymity; it is a systematic project involving institutional coordination, clear responsibilities, and trust construction [3].

2. Technical Logic of Federated Learning in Cross-domain Sharing

2.1 Data Protection Mechanism under a Distributed Collaborative Structure

Federated learning reconstructs the data modeling logic through a distributed collaborative architecture. Its core advantage lies in the design concept of “local computing and global aggregation”, which retains the ownership and control of data to the greatest extent. In federated learning, all participants only need to complete model training locally and upload model parameters or gradients. The central coordination server is only responsible for aggregating updates and does not come into contact with any raw data. This mechanism naturally avoids the reliance of traditional centralized models on data aggregation, and is particularly suitable for scenarios where sensitive information from multiple data sources is inconvenient to share, such as case collaboration among medical institutions and joint risk control models among banks. In addition, as the parameter transmission itself can undergo perturbation processing or encrypted encapsulation, the entire training process has a stronger ability to resist attacks. Typical algorithms such as FedAvg and FedProx not only match the performance of centralized training but also effectively reduce privacy exposure and model bias [4].

2.2 Adaptation of Model Training Strategies among Heterogeneous Participants

In the face of highly heterogeneous data structures in cross-domain environments, federated learning does not rely on a unified data format or sample label, but rather achieves algorithm adaptation among multiple participants through flexible model training paths. Its typical models include horizontal federated learning, vertical federated learning and federated transfer learning: the former is applicable to scenarios where the sample Spaces overlap but the feature distributions are different, such as the integration of diagnosis and treatment data from multiple regional hospitals; The latter is suitable for situations where the features are inconsistent but the individuals overlap, such as when banks and insurance companies model different information about the same customer. In compound scenarios where both samples and features differ, federated transfer learning achieves cross-domain knowledge transfer and local fine-tuning by introducing pre-trained models and domain mapping strategies. It is worth emphasizing that the underlying architecture of federated learning supports local personalization of the model structure and differentiation of parameter scheduling strategies, enabling flexible regulation based on the computing power, data quality and other conditions of the participants [5].

2.3 Analysis of the Integration Models of Mainstream Privacy Protection Mechanisms

The privacy protection capability of federated learning not only stems from its structural advantage of “data localization”, but also is reflected in its ability to embed multiple security mechanisms to build a full-process protection system. In specific applications, mainstream mechanisms such as differential privacy, Secure multi-party Computation (SMPC), and homomorphic encryption can be organically integrated according to the training stage. For instance, local differential privacy can perturb the model update parameters after local training is completed, effectively preventing the server from inverting the original information. SMPC is suitable for encrypting and distributing the intermediate calculation results when aggregating multi-party models, ensuring that “the calculation is available, but the information is invisible”. The homomorphic encryption mechanism allows arithmetic operations to be performed in an encrypted state, enhancing the security of the model upload and aggregation stages. Different mechanisms have their own characteristics in terms of communication complexity, encryption strength, system latency, etc., and the choice should be made by weighing the application scenarios.

3. Optimization Design Path of Privacy-Enhanced Federated Algorithms

3.1 Aggregation Weight Optimization Strategy for Heterogeneous Data Sources

To address the significant differences among multiple participants in terms of data scale, quality, and distribution structure, a dynamic weighted aggregation mechanism can be constructed to enhance the fusion accuracy and stability of the global model [6]. The specific approach is to introduce multi-dimensional indicators to conduct weighted evaluations on each client. Besides the local sample size, the gradient variance, loss degradation rate, and convergence stability of the local model in the most recent multiple rounds of training are also taken into consideration. In actual calculations, a “trust factor” can be set for each client. This factor is dynamically adjusted with each round and is positively correlated with its contribution to model updates. At the same time, establish stability constraints for global parameter aggregation. For clients with drastic changes in trust factors or those that remain in a highly volatile state for a long time, implement gradient suppression or a pause in parameter updates to prevent them from disturbing the training path of the global model. In addition, a “historical validity tracking mechanism” can be introduced on the server side to record the update rate of valid parameters and the prediction accuracy of each node within multiple training cycles, serving as a reference basis for dynamic weight correction. The aggregation function adopts a combination of weighted average and soft gating mechanism to ensure that high-quality updates are given priority for inclusion. At the same time, it sets upload thresholds for low-trust nodes to achieve the regulation and control of training disturbances in heterogeneous data environments.

3.2 Privacy Enhancement Paths for Local Model Perturbation Mechanisms

In the federated learning scenario of cross-domain collaboration, although the local models of the participants do not upload the original data, there is still a risk that the parameter update process will be inferred and restored [7]. To enhance the privacy protection capability of edge nodes, a local differential privacy mechanism can be deployed on the client side to perturbate the model gradient or weight update. The perturbation operation takes the joint strategy of “gradient clipping + noise injection” as its core. Firstly, the gradient threshold for each update within the

training round is set. L2 norm constraints are applied to overly large gradients to suppress privacy exposure caused by abnormal parameter changes. Then, noise is added based on the Laplace or Gaussian distribution to achieve random masking of the target variable. In view of the sensitivity differences of parameters at different levels in the model structure, a “Parameter Importance evaluation module” can be introduced to configure hierarchical perturbations for convolution kernels, embedding layers, output layers, etc., and set higher perturbation intensities for high-sensitivity weights. To ensure the validity of the model after perturbation operations, it is necessary to evaluate the loss variation trend and generalization performance in combination with the local verification mechanism, and dynamically adjust the privacy budget ϵ value during training.

3.3 Fusion Regulation Scheme of the Model Consistency Regularization Mechanism

During the multi-participant collaborative training process, due to the heterogeneous data distribution and the differences in local optimization objectives, federated models often encounter consistency problems such as slow convergence speed and large fluctuations in global performance. To enhance the quality of model fusion, a consistency regularization mechanism can be introduced during the local training stage. The parameter deviation between the local model and the previous round's global model can be used as an additional term of the loss function. Common metrics include L2 norm constraints or KL divergence. This regularization term dynamically participates in the gradient backpass process in each round of updates, guiding the local update direction to maintain consistency with the global trend. In specific training, the regularization coefficient λ can be adaptively adjusted according to the convergence state of the model: in the early stage of training, local learning is the main focus, and the λ weight is relatively low; Gradually increase λ in the middle and later stages of training to guide each client to converge towards a unified goal. To further alleviate the problem of optimization direction deviation caused by structural heterogeneity, a “soft aggregation alignment mechanism” can be constructed. Through the model distillation strategy, the global model is used as the teacher network to provide a consistent output distribution to the client, strengthening semantic layer alignment. Some highly similar nodes can also implement “feature alignment group training” to share the feature mapping structure and improve the efficiency of intra-group aggregation [8].

3.4 Construction of a Federal Security Barrier with Multiple Attack Defense Structures

Although federated learning reduces the risk of raw data exposure through local modeling, it still faces various forms of attacks such as model inversion, member inference, and update poisoning. To enhance system-level protection capabilities, a multi-level federal security barrier can be constructed, with full-chain protection from model input, training process, to parameter aggregation. For model inversion attacks, an adversarial training mechanism can be introduced. During training, a generator can be constructed to simulate the attacker's inference process, and the loss function can be optimized to punish the tendency of information leakage, thereby enhancing the model's resistance to reverse reconstruction. For member inference attacks, a parameter desensitization strategy can be deployed, combined with a sparse gradient upload method, to reduce the sensitivity of a single data point to parameter updates. At the same time, dynamic gradient clipping can be combined to limit the leakage window. To address the issue of malicious nodes uploading contaminated updates, a model verification module can be integrated on the server side. Through projection consistency detection and pre-aggregation parameter review, suspicious updates can be identified to eliminate the source of poisoning. Under the condition of hardware support, it is recommended to build a local training environment based on a Trusted Execution Environment (TEE), combined with a remote audit interface, to achieve traceable management of the training process, parameter calls, and model behavior.

4. Conclusion

The essence of cross-domain data collaboration is not only the integration of technologies but also the reshaping of trust mechanisms and intelligent capabilities. Federated learning, as an important path for distributed intelligent modeling, is not only valuable in “protecting privacy” but also in promoting the efficient flow of data elements under the premise of compliance. The continuous evolution of privacy-enhanced algorithms will drive the modeling logic to shift from “centralized control” to “self-governance”, accelerating the construction of a technical infrastructure for trusted data collaboration. In the future intelligent scenarios where multiple subjects participate, and data is highly sensitive and dense, a federated mechanism that takes into account availability, security, and adaptability will become the key direction of the underlying algorithm architecture.

References

- [1] Zhu H, Huang L, Xie Z. Privacy attack in federated learning is not easy: an experimental study. *Complex Intell Syst.* 2025;11(9):391.
- [2] Yang L, Tong W, Li Z, et al. PECD-DSIIoT: Privacy-enhanced cross-domain data sharing scheme for IIoT. *J Inf Secur Appl.* 2025;93:104128.
- [3] Guan M, Bao H, Wang J, et al. PEFed: Enhancing privacy and efficiency in federated learning via removable perturbation and decentralized encryption. *Inf Fusion.* 2025;122:103187.
- [4] Jiang J, Pei T, Chen J, et al. CDAS: A Secure Cross-Domain Data Sharing Scheme Based on Blockchain. *Information.* 2025;16(5):394.
- [5] Fan H, Fan X, Wei W, et al. Privacy preserving ultra-short-term prediction in clustered wind farms with encrypted data sharing: A secure multi-party computation approach. *Expert Syst Appl.* 2025;278:127218.
- [6] Yang X, Zhang C. A Location Trajectory Privacy Protection Method Based on Generative Adversarial Network and Attention Mechanism. *Comput Mater Contin.* 2024;81(3):3781-804.
- [7] Linwei F, Liming W, Hongjia L. Iterative and mixed-spaces image gradient inversion attack in federated learning. *Cybersecurity.* 2024;7(1).
- [8] Mansoor A, Hadis K, Muhammad T. Integration of Blockchain and Federated Learning for Internet of Things: Recent Advances and Future Challenges. *Comput Secur.* 2021;102355.